

Linear model

All the predictors were significant to the response and a significant regression equation was found:

$$F(4, 7651) = 2.517 \times 10^{-4}, p < 2.2 \times 10^{-16}$$

with

$$R^2 = 0.9294.$$

Gradient boosting (parameter tuning)

According to XGBoost there are several parameters to be tuned. The most important are:

Parameter	Description	Value
eta	After each boosting step, eta shrinks the feature weights to make the boosting process more conservative and less likely to overfit	.1
gamma	Minimum loss reduction required to make a further partition on a leaf node of the tree. A larger gamma implies a more conservative algorithm.	0
max_depth	Maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit.	6
min_child_weight	If the tree partition step results in a leaf node with the sum of instance weight less than min-child-weight, then the building process will give up further partitioning.	2

Chosen by a grid search and 5 fold cross-validation